

Library Sequence Search History

=> fil reg; d que 12

FILE 'REGISTRY' ENTERED AT 16:29:32 ON 05 JUN 2006

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.

COPYRIGHT (C) 2006 American Chemical Society (ACS)

Property values tagged with IC are from the ZIC/VINITI data file provided by InfoChem.

STRUCTURE FILE UPDATES: 4 JUN 2006 HIGHEST RN 886746-35-6

DICTIONARY FILE UPDATES: 4 JUN 2006 HIGHEST RN 886746-35-6

New CAS Information Use Policies, enter HELP USAGETERMS for details.

TSCA INFORMATION NOW CURRENT THROUGH January 6, 2006

Please note that search-term pricing does apply when conducting SmartSELECT searches.

*
* The CA roles and document type information have been removed from *
* the IDE default display format and the ED field has been added, *
* effective March 20, 2005. A new display format, IDERL, is now *
* available and contains the CA role and document type information. *
*

Structure search iteration limits have been increased. See HELP SLIMITS for details.

REGISTRY includes numerically searchable data for experimental and predicted properties as well as tags indicating availability of experimental property data in the original document. For information on property searching in REGISTRY, refer to:

<http://www.cas.org/ONLINE/UG/regprops.html>

L2 1 SEA FILE=REGISTRY ABB=ON KQRTSIRATEGCLPS/SQSFP

=> d sqide 12

L2 ANSWER 1 OF 1 REGISTRY COPYRIGHT 2006 ACS on STN

RN 845509-27-5 REGISTRY

CN L-Serine, L-lysyl-L-glutaminyl-L-arginyl-L-threonyl-L-seryl-L-isoleucyl-L-arginyl-L-alanyl-L-threonyl-L- α -glutamylglycyl-L-cysteinyl-L-leucyl-L-prolyl- (9CI) (CA INDEX NAME)

OTHER NAMES:

CN 12: PN: US20050037972 SEQID: 4 claimed sequence

FS PROTEIN SEQUENCE; STEREOSEARCH

SQL 15

PATENT ANNOTATIONS (PNTE):

Sequence | Patent

Source | Reference

=====+=====

Not Given|US2005037972

Checked
JH
6-7-2006

|claimed
|SEQID 4

SEQ 1 KQTSIRATE GCLPS

=====

HITS AT: 1-15

MF C67 H119 N23 O23 S

SR CA

LC STN Files: CA, CAPLUS, USPATFULL

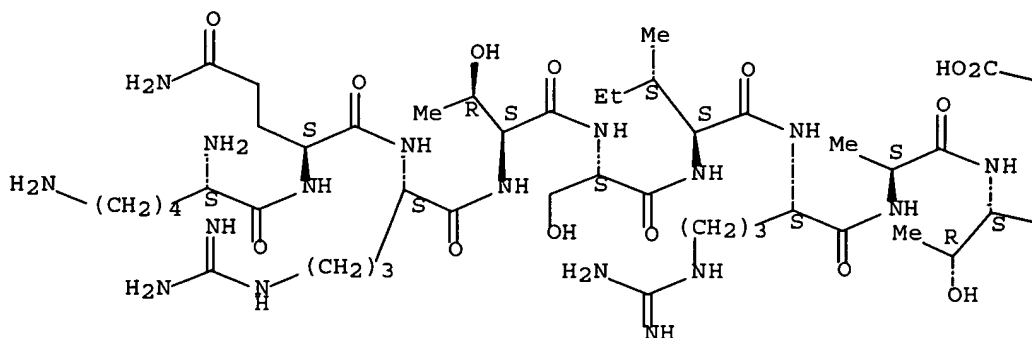
DT.CA Caplus document type: Journal; Patent

RL.P Roles from patents: ANST (Analytical study); BIOL (Biological study);
PREP (Preparation); PRP (Properties); USES (Uses)

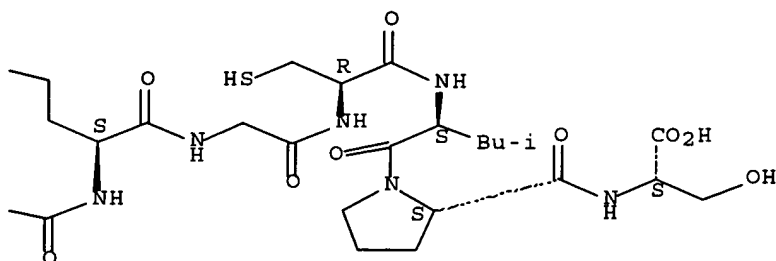
RL.NP Roles from non-patents: BIOL (Biological study)

Absolute stereochemistry.

PAGE 1-A



PAGE 1-B



PROPERTY DATA AVAILABLE IN THE 'PROP' FORMAT

2 REFERENCES IN FILE CA (1907 TO DATE)

2 REFERENCES IN FILE CAPLUS (1907 TO DATE)

=> fil capl uspatf; s l2

FILE 'CAPLUS' ENTERED AT 16:30:04 ON 05 JUN 2006

USE IS SUBJECT TO THE TERMS OF YOUR STN CUSTOMER AGREEMENT.

PLEASE SEE "HELP USAGETERMS" FOR DETAILS.
COPYRIGHT (C) 2006 AMERICAN CHEMICAL SOCIETY (ACS)

FILE 'USPATFULL' ENTERED AT 16:30:04 ON 05 JUN 2006
CA INDEXING COPYRIGHT (C) 2006 AMERICAN CHEMICAL SOCIETY (ACS)

L3 3 L2

=> dup rem l3
PROCESSING COMPLETED FOR L3
L4 2 DUP REM L3 (1 DUPLICATE REMOVED)
ANSWERS '1-2' FROM FILE CAPLUS

=> d ibib ed abs hitrn 1-2; fil hom

L4 ANSWER 1 OF 2 CAPLUS COPYRIGHT 2006 ACS on STN DUPLICATE 1
ACCESSION NUMBER: 2005:140786 CAPLUS Full-text
DOCUMENT NUMBER: 142:233279
TITLE: Phage-display peptides as novel antimicrobial agents
against Haemophilus influenzae, and uses in
identifying bacterial receptors and genes encoding the
same
INVENTOR(S): Bishop-hurley, Sharon L.; Schmidt, Francis J.; Smith,
Arnold L.
PATENT ASSIGNEE(S): The Curators of the University of Missouri, USA
SOURCE: U.S. Pat. Appl. Publ., 24 pp.
CODEN: USXXCO
DOCUMENT TYPE: Patent
LANGUAGE: English
FAMILY ACC. NUM. COUNT: 1
PATENT INFORMATION:

PATENT NO.	KIND	DATE	APPLICATION NO.	DATE
-----	---	----	-----	-----
US 2005037972	A1	20050217	US 2003-655562	20030904
PRIORITY APPLN. INFO.:			US 2002-409909P	P 20020911

ED Entered STN: 18 Feb 2005

AB Whole cell phage-display techniques were used to identify several peptides that bound preferentially to a non-typeable strain of Haemophilus influenzae. These peptides were able to inhibit growth of both H. influenzae and Staphylococcus aureus. Thus, methods for treating bacterial infections, alone or in combination with traditional antibiotics, are envisioned. Also provided is a method for identifying a bacterial receptor comprising (a) providing a sample suspected of comprising a bacterial receptor; (b) providing a peptide comprising the sequence KQRDSRSGYTAPTLV, KKSHHPSSSEWGLNLT, GRHRTSVPTDEVFIT, KQRTSIRATEGCLPS, RNHGTDRAATTIPPLS, VVFLSSRNSAVFTDF, GSRGKHTFVRPTLVF, or FISYSSPSHMGARMR; (c) contacting the sample with the peptide; and (d) identifying a receptor that binds to the peptide. The sample may be a whole bacterium or a bacterial cell wall. The peptide may be fixed to a support, such as a filter, a column, a bead, a dipstick or a gel. The method may further comprise degradative sequencing of said identified bacterial receptor, may further comprise designing a degenerative probe based on the sequence of said identified receptor, may further comprise using the degenerative probe to identify the gene encoding the identified receptor.

IT 845509-27-5P

RL: ARG (Analytical reagent use); BPN (Biosynthetic preparation); DEV (Device component use); PAC (Pharmacological activity); PRP (Properties); THU (Therapeutic use); ANST (Analytical study); BIOL (Biological study); PREP (Preparation); USES (Uses)

(amino acid sequence, antimicrobial peptide; phage-display peptides as novel antimicrobial agents against *Haemophilus influenzae*, and uses in identifying bacterial receptors)

L4 ANSWER 2 OF 2 CAPLUS COPYRIGHT 2006 ACS on STN

ACCESSION NUMBER: 2005:590631 CAPLUS Full-text

DOCUMENT NUMBER: 143:244878

TITLE: Peptides selected for binding to a virulent strain of *Haemophilus influenzae* by phage display are bactericidal

AUTHOR(S): Bishop-Hurley, Sharon L.; Schmidt, Francis J.; Erwin, Alice L.; Smith, Arnold L.

CORPORATE SOURCE: CSIRO Livestock Industries, Rockhampton, 4702, Australia

SOURCE: Antimicrobial Agents and Chemotherapy (2005), 49(7), 2972-2978

CODEN: AMACCQ; ISSN: 0066-4804

PUBLISHER: American Society for Microbiology

DOCUMENT TYPE: Journal

LANGUAGE: English

ED Entered STN: 08 Jul 2005

AB Nontypeable *H. influenzae* (NTHi) is an obligate parasite of the oropharynx of humans, in whom it commonly causes mucosal infections such as otitis media, sinusitis, and bronchitis. We used a subtractive phage display approach to affinity select for peptides binding to the cell surface of a novel invasive NTHi strain R2866 (also called Int1). Over half of the selected phage peptides tested were bactericidal toward R2866 in a dose-dependent manner. Five of the clones encoded the same peptide sequence (KQRTSIRATEGCLPS; clone hi3/17), while the remaining 4 clones encoded unique peptides. All of the bactericidal phage peptides but one were cationic and had similar phys.-chemical properties. Clone hi3/17 possessed a similar level of activity toward a panel of clin. NTHi isolates and *H. influenzae* type b strains but lacked bactericidal activity toward gram-pos. (*Enterococcus faecalis*, *Staphylococcus aureus*) and gram-neg. (*Proteus mirabilis*, *Pseudomonas aeruginosa*, and *Salmonella enterica*) bacteria. These data indicate that peptides binding to bacterial surface structures isolated by phage display may prove of value in developing new antibiotics.

IT 845509-27-5

RL: BSU (Biological study, unclassified); BIOL (Biological study)
(peptides binding to a virulent strain of *Haemophilus influenzae* by phage display are bactericidal)

REFERENCE COUNT: 37 THERE ARE 37 CITED REFERENCES AVAILABLE FOR THIS RECORD. ALL CITATIONS AVAILABLE IN THE RE FORMAT

FILE 'HOME' ENTERED AT 16:30:17 ON 05 JUN 2006

=>

GenCore version 5.1.9
 Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2006, 05:12:21 ; Search time 197 Seconds
 (without alignments)
 34.813 Million cell updates/sec

Title: US-10-655-562A-4
 Perfect score: 77
 Sequence: 1 KQRTSIRATEGCLPS 15

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 2589679 seqs, 457216429 residues

Total number of hits satisfying chosen parameters: 2589679

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : A_Geneseq_8:*
 1: geneseqp1980s:*
 2: geneseqp1990s:*
 3: geneseqp2000s:*
 4: geneseqp2001s:*
 5: geneseqp2002s:*
 6: geneseqp2003as:*
 7: geneseqp2003bs:*
 8: geneseqp2004s:*
 9: geneseqp2005s:*
 10: geneseqp2006s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	% Query		Length	DB	ID	Description
	Score	Match				
1	77	100.0	15	9	ADY21128	Ady21128 Haemophil
2	42	54.5	109	4	AAM96603	Aam96603 Human rep
3	42	54.5	109	4	ABB96571	Abb96571 Human tes
4	42	54.5	249	2	AAW79090	Aaw79090 Human sec
5	42	54.5	249	4	AAU12226	Aau12226 Human PRO
6	42	54.5	249	4	AAU09180	Aau09180 Human PRO
7	42	54.5	249	5	AAM49039	Aam49039 Human tes

8	42	54.5	249	5	ABP61791	Abp61791	Human	pol
9	42	54.5	249	6	ABO17670	Abo17670	Novel	hum
10	42	54.5	249	6	ABU80924	Abu80924	Human	PRO
11	42	54.5	249	6	ABU66624	Abu66624	Human	PRO
12	42	54.5	249	6	ABU59705	Abu59705	Novel	sec
13	42	54.5	249	6	ABO24895	Abo24895	Human	sec
14	42	54.5	249	6	ABU66900	Abu66900	Human	sec
15	42	54.5	249	6	ABP54449	Abp54449	Putative	
16	42	54.5	249	6	ADA45629	Ada45629	Novel	hum
17	42	54.5	249	6	ADA76060	Ada76060	Human	PRO
18	42	54.5	249	6	ADA18710	Ada18710	Human	PRO
19	42	54.5	249	6	ADA61333	Ada61333	Homo sapi	
20	42	54.5	249	6	ADB19118	Adb19118	Novel	hum
21	42	54.5	249	6	ADB27659	Adb27659	Human	PRO
22	42	54.5	249	6	ADA86138	Ada86138	Novel	hum
23	42	54.5	249	6	ADB15702	Adb15702	Human	PRO
24	42	54.5	249	6	ADA47488	Ada47488	Human	PRO
25	42	54.5	249	6	ADA67283	Ada67283	Human	PRO
26	42	54.5	249	6	ADB30290	Adb30290	Human	PRO
27	42	54.5	249	6	ADA85586	Ada85586	Novel	hum
28	42	54.5	249	6	ADA96798	Ada96798	Human	PRO
29	42	54.5	249	6	ADA79102	Ada79102	Human	PRO
30	42	54.5	249	6	ADA87241	Ada87241	Novel	hum
31	42	54.5	249	6	ADB16443	Adb16443	Human	PRO
32	42	54.5	249	6	ADA91535	Ada91535	Novel	hum
33	42	54.5	249	6	ADB14598	Adb14598	Human	PRO
34	42	54.5	249	6	ADB18559	Adb18559	Novel	hum
35	42	54.5	249	6	ADA93774	Ada93774	Human	PRO
36	42	54.5	249	6	ADB19670	Adb19670	Novel	hum
37	42	54.5	249	6	ADB12982	Adb12982	Human	PRO
38	42	54.5	249	6	ABO43203	Abo43203	Novel	hum
39	42	54.5	249	6	ADA74236	Ada74236	Human	PRO
40	42	54.5	249	6	ADB24469	Adb24469	Human	PRO
41	42	54.5	249	6	ADA81993	Ada81993	Human	PRO
42	42	54.5	249	6	ADA74956	Ada74956	Human	PRO
43	42	54.5	249	6	ADA85034	Ada85034	Novel	hum
44	42	54.5	249	6	ADA84482	Ada84482	Novel	hum
45	42	54.5	249	6	ADB29738	Adb29738	Human	PRO

ALIGNMENTS

RESULT 1

ADY21128

ID ADY21128 standard; peptide; 15 AA.

XX

AC ADY21128;

XX

DT 05-MAY-2005 (first entry)

XX

DE Haemophilus influenza-binding phage display method peptide #14.

XX

KW antibacterial; phage display; protein interaction;

KW haemophilus influenzae infection; staphylococcus aureus infection.

XX

OS Unidentified.

XX
 PN US2005037972-A1.
 XX
 PD 17-FEB-2005.
 XX
 PF 04-SEP-2003; 2003US-00655562.
 XX
 PR 11-SEP-2002; 2002US-0409909P.
 XX
 PA (UMOR) UNIV MISSOURI.
 XX
 PI Bishop-Hurley SL, Schmidt FJ, Smith AL;
 XX
 DR WPI; 2005-172291/18.
 XX
 PT Novel isolated peptide derived from Haemophilus influenzae, useful for
 PT inhibiting growth of Staphylococcal or Haemophilus species such as
 PT Staphylococcus aureus or H.influenzae, and treating/preventing bacterial
 PT infection in subject.
 XX
 PS Claim 1; Page 15; 24pp; English.
 XX
 CC The invention relates to an isolated peptide (I) derived from Haemophilus
 CC influenzae, and comprising 15-50 residues of any one of 8 fully defined
 CC sequences given in specification. (I) is useful for inhibiting the growth
 CC of a Staphylococcal or Haemophilus sp. such as Staphylococcus aureus or
 CC H.influenzae . The peptide is 15-50 residues, preferably 15 residues in
 CC length. The method involves contacting the species with (I), and
 CC contacting the species with a chemopharmaceutical antibiotic. (I) is
 CC useful for treating or preventing a bacterial infection in a subject,
 CC which involves contacting the subject with (I), to inhibit the growth of
 CC bacteria in vivo . (I) is useful for preventing bacterial growth in a
 CC solution or bacterial attachment or growth on an abiotic surface, which
 CC involves mixing the solution with (I) or coating the abiotic surface with
 CC (I) to inhibit the growth of bacteria in vivo . The surface is part of a
 CC medical device. (I) is useful for identifying a bacterial receptor in a
 CC sample, which involves providing a sample suspected of comprising a
 CC bacterial receptor, contacting the sample with (I), and identifying a
 CC receptor that binds to (I). The sample is a whole bacterium or bacterial
 CC cell wall. (I) is fixed to a support such as a filter, column, bead,
 CC dipstick or gel. The method further involves degradative sequencing of
 CC the identified receptor, designing a degenerative probe based on the
 CC sequence of the identified receptor and using the degenerative probe to
 CC identify the gene encoding the identified receptor. (Note: this sequence
 CC is given as SEQ ID NO: 4 in the claims of the patent but does not
 CC corresponds to the sequence given as SEQ ID NO: 4 in the Sequence Listing
 CC of the specification).
 XX
 SQ Sequence 15 AA;

Query Match 100.0%; Score 77; DB 9; Length 15;
 Best Local Similarity 100.0%; Pred. No. 2.8e-06;
 Matches 15; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 KQRTSIRATEGCLPS 15
 |||||
 Db 1 KQRTSIRATEGCLPS 15

RESULT 2

AAM96603

ID AAM96603 standard; protein; 109 AA.

XX

AC AAM96603;

XX

DT 21-NOV-2001 (first entry)

XX

DE Human reproductive system related antigen SEQ ID NO: 5261.

XX

KW Human; reproductive system related antigen; reproductive system disorder;
KW cancer; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200155320-A2.

XX

PD 02-AUG-2001.

XX

PF 17-JAN-2001; 2001WO-US001339.

XX

PR 31-JAN-2000; 2000US-0179065P.

PR 04-FEB-2000; 2000US-0180628P.

PR 24-FEB-2000; 2000US-0184664P.

PR 02-MAR-2000; 2000US-0186350P.

PR 16-MAR-2000; 2000US-0189874P.

PR 17-MAR-2000; 2000US-0190076P.

PR 18-APR-2000; 2000US-0198123P.

PR 19-MAY-2000; 2000US-0205515P.

PR 07-JUN-2000; 2000US-0209467P.

PR 28-JUN-2000; 2000US-0214886P.

PR 30-JUN-2000; 2000US-0215135P.

PR 07-JUL-2000; 2000US-0216647P.

PR 07-JUL-2000; 2000US-0216880P.

PR 11-JUL-2000; 2000US-0217487P.

PR 11-JUL-2000; 2000US-0217496P.

PR 14-JUL-2000; 2000US-0218290P.

PR 26-JUL-2000; 2000US-0220963P.

PR 26-JUL-2000; 2000US-0220964P.

PR 14-AUG-2000; 2000US-0224518P.

PR 14-AUG-2000; 2000US-0224519P.

PR 14-AUG-2000; 2000US-0225213P.

PR 14-AUG-2000; 2000US-0225214P.

PR 14-AUG-2000; 2000US-0225266P.

PR 14-AUG-2000; 2000US-0225267P.

PR 14-AUG-2000; 2000US-0225268P.

PR 14-AUG-2000; 2000US-0225270P.

PR 14-AUG-2000; 2000US-0225447P.

PR 14-AUG-2000; 2000US-0225757P.

PR 14-AUG-2000; 2000US-0225758P.

PR 14-AUG-2000; 2000US-0225759P.

PR 18-AUG-2000; 2000US-0226279P.

PR 22-AUG-2000; 2000US-0226681P.

PR 22-AUG-2000; 2000US-0226868P.

PR 22-AUG-2000; 2000US-0227182P.

PR 23-AUG-2000; 2000US-0227009P.
PR 30-AUG-2000; 2000US-0228924P.
PR 01-SEP-2000; 2000US-0229287P.
PR 01-SEP-2000; 2000US-0229343P.
PR 01-SEP-2000; 2000US-0229344P.
PR 01-SEP-2000; 2000US-0229345P.
PR 05-SEP-2000; 2000US-0229509P.
PR 05-SEP-2000; 2000US-0229513P.
PR 06-SEP-2000; 2000US-0230437P.
PR 06-SEP-2000; 2000US-0230438P.
PR 08-SEP-2000; 2000US-0231242P.
PR 08-SEP-2000; 2000US-0231243P.
PR 08-SEP-2000; 2000US-0231244P.
PR 08-SEP-2000; 2000US-0231413P.
PR 08-SEP-2000; 2000US-0231414P.
PR 08-SEP-2000; 2000US-0232080P.
PR 08-SEP-2000; 2000US-0232081P.
PR 12-SEP-2000; 2000US-0231968P.
PR 14-SEP-2000; 2000US-0232397P.
PR 14-SEP-2000; 2000US-0232398P.
PR 14-SEP-2000; 2000US-0232399P.
PR 14-SEP-2000; 2000US-0232400P.
PR 14-SEP-2000; 2000US-0232401P.
PR 14-SEP-2000; 2000US-0233063P.
PR 14-SEP-2000; 2000US-0233064P.
PR 14-SEP-2000; 2000US-0233065P.
PR 21-SEP-2000; 2000US-0234223P.
PR 21-SEP-2000; 2000US-0234274P.
PR 25-SEP-2000; 2000US-0234997P.
PR 25-SEP-2000; 2000US-0234998P.
PR 26-SEP-2000; 2000US-0235484P.
PR 27-SEP-2000; 2000US-0235834P.
PR 27-SEP-2000; 2000US-0235836P.
PR 29-SEP-2000; 2000US-0236327P.
PR 29-SEP-2000; 2000US-0236367P.
PR 29-SEP-2000; 2000US-0236368P.
PR 29-SEP-2000; 2000US-0236369P.
PR 29-SEP-2000; 2000US-0236370P.
PR 02-OCT-2000; 2000US-0236802P.
PR 02-OCT-2000; 2000US-0237037P.
PR 02-OCT-2000; 2000US-0237038P.
PR 02-OCT-2000; 2000US-0237039P.
PR 02-OCT-2000; 2000US-0237040P.
PR 13-OCT-2000; 2000US-0239935P.
PR 13-OCT-2000; 2000US-0239937P.
PR 20-OCT-2000; 2000US-0240960P.
PR 20-OCT-2000; 2000US-0241221P.
PR 20-OCT-2000; 2000US-0241785P.
PR 20-OCT-2000; 2000US-0241786P.
PR 20-OCT-2000; 2000US-0241787P.
PR 20-OCT-2000; 2000US-0241808P.
PR 20-OCT-2000; 2000US-0241809P.
PR 20-OCT-2000; 2000US-0241826P.
PR 01-NOV-2000; 2000US-0244617P.
PR 08-NOV-2000; 2000US-0246474P.
PR 08-NOV-2000; 2000US-0246475P.
PR 08-NOV-2000; 2000US-0246476P.

PR 08-NOV-2000; 2000US-0246477P.
PR 08-NOV-2000; 2000US-0246478P.
PR 08-NOV-2000; 2000US-0246523P.
PR 08-NOV-2000; 2000US-0246524P.
PR 08-NOV-2000; 2000US-0246525P.
PR 08-NOV-2000; 2000US-0246526P.
PR 08-NOV-2000; 2000US-0246527P.
PR 08-NOV-2000; 2000US-0246528P.
PR 08-NOV-2000; 2000US-0246532P.
PR 08-NOV-2000; 2000US-0246609P.
PR 08-NOV-2000; 2000US-0246610P.
PR 08-NOV-2000; 2000US-0246611P.
PR 08-NOV-2000; 2000US-0246613P.
PR 17-NOV-2000; 2000US-0249207P.
PR 17-NOV-2000; 2000US-0249208P.
PR 17-NOV-2000; 2000US-0249209P.
PR 17-NOV-2000; 2000US-0249210P.
PR 17-NOV-2000; 2000US-0249211P.
PR 17-NOV-2000; 2000US-0249212P.
PR 17-NOV-2000; 2000US-0249213P.
PR 17-NOV-2000; 2000US-0249214P.
PR 17-NOV-2000; 2000US-0249215P.
PR 17-NOV-2000; 2000US-0249216P.
PR 17-NOV-2000; 2000US-0249217P.
PR 17-NOV-2000; 2000US-0249218P.
PR 17-NOV-2000; 2000US-0249244P.
PR 17-NOV-2000; 2000US-0249245P.
PR 17-NOV-2000; 2000US-0249264P.
PR 17-NOV-2000; 2000US-0249265P.
PR 17-NOV-2000; 2000US-0249297P.
PR 17-NOV-2000; 2000US-0249299P.
PR 17-NOV-2000; 2000US-0249300P.
PR 01-DEC-2000; 2000US-0250160P.
PR 01-DEC-2000; 2000US-0250391P.
PR 05-DEC-2000; 2000US-0251030P.
PR 05-DEC-2000; 2000US-0251988P.
PR 05-DEC-2000; 2000US-0256719P.
PR 06-DEC-2000; 2000US-0251479P.
PR 08-DEC-2000; 2000US-0251856P.
PR 08-DEC-2000; 2000US-0251868P.
PR 08-DEC-2000; 2000US-0251869P.
PR 08-DEC-2000; 2000US-0251989P.
PR 08-DEC-2000; 2000US-0251990P.
PR 11-DEC-2000; 2000US-0254097P.
PR 05-JAN-2001; 2001US-0259678P.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Rosen CA, Barash SC, Ruben SM;

XX

DR WPI; 2001-465570/50.

DR N-PSDB; AAL02573.

XX

PT Isolated nucleic acid molecule encoding a reproductive system antigen is
PT used in preventing, treating or ameliorating a medical condition.

XX

PS Claim 11; SEQ ID NO 5261; 1297pp + Sequence Listing; English.

XX
CC The present invention provides the protein and coding sequences of a
CC number of human reproductive system related antigens. These can be used
CC in the prevention and treatment of reproductive system disorders,
CC including cancer. The present sequence is a protein of the invention
XX
SQ Sequence 109 AA;

Query Match 54.5%; Score 42; DB 4; Length 109;
Best Local Similarity 63.6%; Pred. No. 30;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TSIRATEGCLP 14
|:| |:| |:|
Db 94 TAILATKGCIP 104

Search completed: June 6, 2006, 05:15:55
Job time : 200 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2006, 05:21:10 ; Search time 50 Seconds
(without alignments)
26.259 Million cell updates/sec

Title: US-10-655-562A-4
Perfect score: 77
Sequence: 1 KQRTSIRATEGCLPS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 650591 seqs, 87530628 residues

Total number of hits satisfying chosen parameters: 650591

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /EMC_Celerra_SIDS3/ptodata/2/iaa/5_COMB.pep:*
2: /EMC_Celerra_SIDS3/ptodata/2/iaa/6_COMB.pep:*
3: /EMC_Celerra_SIDS3/ptodata/2/iaa/7_COMB.pep:*
4: /EMC_Celerra_SIDS3/ptodata/2/iaa/H_COMB.pep:*
5: /EMC_Celerra_SIDS3/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /EMC_Celerra_SIDS3/ptodata/2/iaa/RE_COMB.pep:*
7: /EMC_Celerra_SIDS3/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result		%					
No.	Score	Query Match	Length	DB	ID	Description	
1	42	54.5	55	2	US-09-621-976-7108	Sequence 7108, Ap	
2	42	54.5	88	2	US-09-621-976-5239	Sequence 5239, Ap	
3	42	54.5	120	2	US-09-621-976-5688	Sequence 5688, Ap	
4	42	54.5	450	2	US-09-252-991A-26556	Sequence 26556, A	
5	42	54.5	854	2	US-09-206-551-16	Sequence 16, Appl	
6	41	53.2	94	2	US-09-489-039A-8163	Sequence 8163, Ap	
7	39	50.6	161	2	US-09-270-767-33980	Sequence 33980, A	
8	39	50.6	161	2	US-09-270-767-49197	Sequence 49197, A	
9	39	50.6	268	2	US-09-252-991A-31279	Sequence 31279, A	
10	39	50.6	291	2	US-09-252-991A-19371	Sequence 19371, A	
11	39	50.6	363	2	US-09-205-258-553	Sequence 553, App	
12	39	50.6	363	2	US-10-004-860-553	Sequence 553, App	
13	39	50.6	624	2	US-09-252-991A-23659	Sequence 23659, A	
14	38	49.4	165	2	US-09-252-991A-17601	Sequence 17601, A	
15	37	48.1	72	2	US-08-469-260A-453	Sequence 453, App	
16	37	48.1	72	2	US-08-488-446-453	Sequence 453, App	
17	37	48.1	72	2	US-08-467-344A-453	Sequence 453, App	
18	37	48.1	72	2	US-08-424-550B-453	Sequence 453, App	
19	37	48.1	91	2	US-09-376-781-2	Sequence 2, Appli	
20	37	48.1	135	2	US-09-252-991A-22855	Sequence 22855, A	
21	37	48.1	532	2	US-09-533-427-6	Sequence 6, Appli	
22	37	48.1	532	2	US-09-717-789C-6	Sequence 6, Appli	
23	37	48.1	588	2	US-09-533-427-5	Sequence 5, Appli	
24	37	48.1	588	2	US-09-717-789C-5	Sequence 5, Appli	
25	37	48.1	724	2	US-09-533-427-4	Sequence 4, Appli	
26	37	48.1	724	2	US-09-717-789C-4	Sequence 4, Appli	
27	36	46.8	78	2	US-09-621-976-5240	Sequence 5240, Ap	
28	36	46.8	192	2	US-09-252-991A-27287	Sequence 27287, A	
29	36	46.8	354	2	US-09-134-000C-3663	Sequence 3663, Ap	
30	36	46.8	361	1	US-08-415-751-7	Sequence 7, Appli	
31	36	46.8	363	2	US-09-252-991A-30821	Sequence 30821, A	
32	36	46.8	393	2	US-09-248-796A-16799	Sequence 16799, A	
33	36	46.8	505	2	US-09-252-991A-23615	Sequence 23615, A	
34	36	46.8	605	2	US-09-252-991A-19462	Sequence 19462, A	
35	36	46.8	1278	2	US-09-462-136-2	Sequence 2, Appli	
36	36	46.8	1318	2	US-09-949-016-10152	Sequence 10152, A	
37	35	45.5	22	2	US-09-205-258-884	Sequence 884, App	
38	35	45.5	22	2	US-10-004-860-884	Sequence 884, App	
39	35	45.5	59	1	US-08-018-129-7	Sequence 7, Appli	
40	35	45.5	59	1	US-08-448-250-7	Sequence 7, Appli	
41	35	45.5	59	2	US-09-282-257-7	Sequence 7, Appli	
42	35	45.5	83	2	US-09-674-973A-344	Sequence 344, App	
43	35	45.5	84	2	US-09-674-973A-343	Sequence 343, App	
44	35	45.5	91	2	US-09-674-973A-348	Sequence 348, App	
45	35	45.5	92	2	US-09-674-973A-347	Sequence 347, App	

ALIGNMENTS

RESULT 1

US-09-621-976-7108

; Sequence 7108, Application US/09621976
; Patent No. 6639063
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Jobert, S.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.
; FILE REFERENCE: GENSET.054PR2
; CURRENT APPLICATION NUMBER: US/09/621,976
; CURRENT FILING DATE: 2000-07-21
; NUMBER OF SEQ ID NOS: 19335
; SOFTWARE: Patent.pm
; SEQ ID NO 7108
; LENGTH: 55
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: 27
; OTHER INFORMATION: Xaa = Cys,Phe
; NAME/KEY: UNSURE
; LOCATION: 53
; OTHER INFORMATION: Xaa = Met,Arg

US-09-621-976-7108

Query Match 54.5%; Score 42; DB 2; Length 55;
Best Local Similarity 63.6%; Pred. No. 2.9;
Matches 7; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 4 TSIRATEGCLP 14
|:| ||:|:|
Db 40 TAILATKGCIP 50

Search completed: June 6, 2006, 05:22:34
Job time : 51 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2006, 05:33:06 ; Search time 184 Seconds
(without alignments)
37.762 Million cell updates/sec

Title: US-10-655-562A-4
Perfect score: 77
Sequence: 1 KQRTSIRATEGCLPS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2097797 seqs, 463214858 residues

Total number of hits satisfying chosen parameters: 2097797

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:*

- 1: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
- 2: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
- 3: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
- 4: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
- 5: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
- 6: /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		Length	DB	ID	Description
		Match					
1	44	57.1		78	4	US-10-437-963-115135	Sequence 115135,
2	43	55.8		80	4	US-10-425-115-215070	Sequence 215070,
3	42	54.5		109	3	US-09-764-891-5261	Sequence 5261, Ap
4	42	54.5		117	4	US-10-437-963-135764	Sequence 135764,
5	42	54.5		208	4	US-10-425-115-276978	Sequence 276978,
6	42	54.5		249	3	US-09-745-763-9	Sequence 9, Appli
7	42	54.5		249	4	US-10-028-072-110	Sequence 110, App
8	42	54.5		249	4	US-10-140-808-110	Sequence 110, App
9	42	54.5		249	4	US-10-121-049-110	Sequence 110, App
10	42	54.5		249	4	US-10-123-904-110	Sequence 110, App
11	42	54.5		249	4	US-10-140-470-110	Sequence 110, App
12	42	54.5		249	4	US-10-175-746-110	Sequence 110, App
13	42	54.5		249	4	US-10-176-918-110	Sequence 110, App
14	42	54.5		249	4	US-10-176-921-110	Sequence 110, App
15	42	54.5		249	4	US-10-137-865-110	Sequence 110, App
16	42	54.5		249	4	US-10-140-474-110	Sequence 110, App
17	42	54.5		249	4	US-10-142-431-110	Sequence 110, App
18	42	54.5		249	4	US-10-143-114-110	Sequence 110, App
19	42	54.5		249	4	US-10-142-419-110	Sequence 110, App
20	42	54.5		249	4	US-10-123-262-110	Sequence 110, App
21	42	54.5		249	4	US-10-142-423-110	Sequence 110, App
22	42	54.5		249	4	US-10-121-050-110	Sequence 110, App
23	42	54.5		249	4	US-10-141-755-110	Sequence 110, App
24	42	54.5		249	4	US-10-143-032-110	Sequence 110, App
25	42	54.5		249	4	US-10-123-108-110	Sequence 110, App
26	42	54.5		249	4	US-10-123-236-110	Sequence 110, App

27	42	54.5	249	4	US-10-123-261-110	Sequence 110, App
28	42	54.5	249	4	US-10-140-921-110	Sequence 110, App
29	42	54.5	249	4	US-10-140-928-110	Sequence 110, App
30	42	54.5	249	4	US-10-121-045-110	Sequence 110, App
31	42	54.5	249	4	US-10-123-292-110	Sequence 110, App
32	42	54.5	249	4	US-10-123-903-110	Sequence 110, App
33	42	54.5	249	4	US-10-124-819-110	Sequence 110, App
34	42	54.5	249	4	US-10-124-822-110	Sequence 110, App
35	42	54.5	249	4	US-10-140-925-110	Sequence 110, App
36	42	54.5	249	4	US-10-160-498-110	Sequence 110, App
37	42	54.5	249	4	US-10-124-824-110	Sequence 110, App
38	42	54.5	249	4	US-10-127-825A-110	Sequence 110, App
39	42	54.5	249	4	US-10-127-829A-110	Sequence 110, App
40	42	54.5	249	4	US-10-127-835A-110	Sequence 110, App
41	42	54.5	249	4	US-10-127-839A-110	Sequence 110, App
42	42	54.5	249	4	US-10-127-901A-110	Sequence 110, App
43	42	54.5	249	4	US-10-128-693A-110	Sequence 110, App
44	42	54.5	249	4	US-10-131-813A-110	Sequence 110, App
45	42	54.5	249	4	US-10-131-818A-110	Sequence 110, App

ALIGNMENTS

RESULT 1

US-10-437-963-115135
 ; Sequence 115135, Application US/10437963
 ; Publication No. US20040123343A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kovalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Wu, Wei
 ; APPLICANT: Boukharov, Andrey A.
 ; APPLICANT: Barbazuk, Brad
 ; APPLICANT: Li, Ping
 ; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules
 Associated With
 ; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
 ; FILE REFERENCE: 38-21(53221)B
 ; CURRENT APPLICATION NUMBER: US/10/437,963
 ; CURRENT FILING DATE: 2003-05-14
 ; NUMBER OF SEQ ID NOS: 204966
 ; SEQ ID NO 115135
 ; LENGTH: 78
 ; TYPE: PRT
 ; ORGANISM: Oryza sativa
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: PAT_MRT4530_18760C.1.pep
 US-10-437-963-115135

Query Match 57.1%; Score 44; DB 4; Length 78;
 Best Local Similarity 57.1%; Pred. No. 10;
 Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 1 KQRTSIRATEGCLP 14

Db 53 KORSSRKGEGCFP 66

Search completed: June 6, 2006, 05:36:25
Job time : 185 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

```
Run on:      June  6, 2006, 05:33:25 ; Search time 15 Seconds
              (without alignments)
              11.565 Million cell updates/sec
```

```
Title:      US-10-655-562A-4
Perfect score:  77
Sequence:      1 KQRTSIRATEGCLPS 15
```

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 58871 seqs, 11565156 residues

Total number of hits satisfying chosen parameters: 58871

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

```
Database :      Published_Applications_AA_New:*
1:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
2:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
3:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
4:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
5:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
6:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
7:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
8:   /EMC_Celerra_SIDS3/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		Length	DB	ID	Description
		Query	Match				
1	39	50.6		221	7	US-11-293-697-2989	Sequence 2989, Ap
2	37	48.1		324	6	US-10-953-349-22573	Sequence 22573, A

3	37	48.1	439	6	US-10-953-349-15731	Sequence 15731, A
4	37	48.1	452	6	US-10-953-349-15730	Sequence 15730, A
5	36	46.8	151	7	US-11-293-697-3091	Sequence 3091, Ap
6	36	46.8	234	6	US-10-953-349-25569	Sequence 25569, A
7	36	46.8	432	7	US-11-293-697-3406	Sequence 3406, Ap
8	35	45.5	364	6	US-10-953-349-10990	Sequence 10990, A
9	35	45.5	529	7	US-11-312-958-26	Sequence 26, Appl
10	35	45.5	945	7	US-11-293-697-2855	Sequence 2855, Ap
11	35	45.5	945	7	US-11-293-697-3079	Sequence 3079, Ap
12	34	44.2	236	7	US-11-293-697-4829	Sequence 4829, Ap
13	34	44.2	338	6	US-10-953-349-2039	Sequence 2039, Ap
14	34	44.2	340	6	US-10-953-349-2038	Sequence 2038, Ap
15	34	44.2	769	6	US-10-511-937-3015	Sequence 3015, Ap
16	34	44.2	813	7	US-11-293-697-3901	Sequence 3901, Ap
17	33.5	43.5	247	6	US-10-953-349-16451	Sequence 16451, A
18	33.5	43.5	258	6	US-10-953-349-16450	Sequence 16450, A
19	33.5	43.5	319	6	US-10-953-349-16449	Sequence 16449, A
20	33	42.9	157	6	US-10-953-349-27909	Sequence 27909, A
21	33	42.9	233	6	US-10-953-349-16422	Sequence 16422, A
22	33	42.9	295	7	US-11-242-111-24	Sequence 24, Appl
23	33	42.9	432	6	US-10-196-749-74	Sequence 74, Appl
24	33	42.9	562	6	US-10-953-349-20235	Sequence 20235, A
25	33	42.9	568	6	US-10-953-349-20234	Sequence 20234, A
26	33	42.9	599	6	US-10-953-349-20233	Sequence 20233, A
27	32	41.6	134	6	US-10-953-349-38480	Sequence 38480, A
28	32	41.6	188	6	US-10-953-349-36581	Sequence 36581, A
29	32	41.6	242	6	US-10-953-349-28679	Sequence 28679, A
30	32	41.6	288	6	US-10-953-349-28678	Sequence 28678, A
31	32	41.6	321	7	US-11-140-450-66	Sequence 66, Appl
32	32	41.6	331	6	US-10-953-349-28677	Sequence 28677, A
33	32	41.6	369	6	US-10-953-349-7986	Sequence 7986, Ap
34	32	41.6	477	6	US-10-505-928-515	Sequence 515, App
35	32	41.6	516	6	US-10-953-349-7985	Sequence 7985, Ap
36	32	41.6	547	6	US-10-953-349-7984	Sequence 7984, Ap
37	32	41.6	574	7	US-11-293-697-2802	Sequence 2802, Ap
38	32	41.6	1043	7	US-11-293-697-3097	Sequence 3097, Ap
39	32	41.6	1336	7	US-11-106-014-92	Sequence 92, Appl
40	31.5	40.9	196	7	US-11-293-697-3347	Sequence 3347, Ap
41	31	40.3	102	7	US-11-293-697-4024	Sequence 4024, Ap
42	31	40.3	115	6	US-10-953-349-10173	Sequence 10173, A
43	31	40.3	115	7	US-11-293-697-3614	Sequence 3614, Ap
44	31	40.3	117	7	US-11-293-697-2985	Sequence 2985, Ap
45	31	40.3	123	6	US-10-953-349-40076	Sequence 40076, A

ALIGNMENTS

RESULT 1

US-11-293-697-2989

; Sequence 2989, Application US/11293697

; Publication No. US20060105376A1

; GENERAL INFORMATION:

; APPLICANT: HELIX RESEARCH INSTITUTE

; TITLE OF INVENTION: Novel full length cDNA

; FILE REFERENCE: H1-A0106

; CURRENT APPLICATION NUMBER: US/11/293,697

; CURRENT FILING DATE: 2005-12-05
; PRIOR APPLICATION NUMBER: US/10/108,260
; PRIOR FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 5458
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2989
; LENGTH: 221
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-293-697-2989

Query Match 50.6%; Score 39; DB 7; Length 221;
Best Local Similarity 57.1%; Pred. No. 3;
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 QRTSIRATEGCLPS 15
: || : || ||||
Db 175 KRTPLCATAPCLPS 188

Search completed: June 6, 2006, 05:36:46
Job time : 15 secs

GenCore version 5.1.9
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2006, 05:16:10 ; Search time 38 Seconds
(without alignments)
37.980 Million cell updates/sec

Title: US-10-655-562A-4
Perfect score: 77
Sequence: 1 KQRTSIRATEGCLPS 15

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	45	58.4	244	2	S48492	DCG1 protein - yea
2	42	54.5	854	1	VCLJSI	env polyprotein pr
3	40.5	52.6	312	2	C71136	hypothetical prote
4	40	51.9	290	2	F81700	DNA polymerase III
5	39	50.6	160	2	S56204	probable membrane
6	39	50.6	331	2	G69200	conserved hypothet
7	38	49.4	191	2	T49232	hypothetical prote
8	38	49.4	968	2	C82452	hypothetical prote
9	38	49.4	1220	2	T32916	hypothetical prote
10	37	48.1	188	2	T20235	hypothetical prote
11	37	48.1	340	2	T25919	hypothetical prote
12	37	48.1	359	2	T21247	hypothetical prote
13	37	48.1	487	2	F70765	hypothetical prote
14	37	48.1	642	2	D64491	hypothetical prote
15	37	48.1	1071	2	D86279	hypothetical prote
16	36	46.8	113	2	AH2677	hypothetical prote
17	36	46.8	169	2	C87610	conserved hypothet
18	36	46.8	206	2	T16153	hypothetical prote
19	36	46.8	275	2	B97323	multidrug-efflux t
20	36	46.8	308	2	T24912	hypothetical prote
21	36	46.8	325	2	F83503	hypothetical prote
22	36	46.8	346	2	F70666	probable alcohol d
23	36	46.8	482	2	H86447	hypothetical prote
24	36	46.8	665	2	H87468	ubiquinol oxidase
25	36	46.8	993	2	C55226	cylM protein - Ent
26	36	46.8	2351	2	G71415	hypothetical prote
27	36	46.8	2567	2	A49551	filamin, Muller ce
28	36	46.8	4006	2	T09070	probable tenascin
29	35.5	46.1	644	2	T24366	hypothetical prote
30	35.5	46.1	679	2	T24365	hypothetical prote
31	35	45.5	147	2	A75196	hypothetical prote
32	35	45.5	189	2	T19559	hypothetical prote
33	35	45.5	273	2	E95095	hypothetical prote
34	35	45.5	287	2	C75635	phosphoenolpyruvat
35	35	45.5	335	2	S25212	prsG protein - Esc
36	35	45.5	335	2	S25229	G-minor fimbrial p
37	35	45.5	375	2	H97560	alcohol dehydrogen
38	35	45.5	375	2	AH2781	alcohol dehydrogen
39	35	45.5	384	2	S25771	gas1 protein - mou
40	35	45.5	409	2	E91246	probable L-sorbose
41	35	45.5	413	2	B86094	probable L-sorbose
42	35	45.5	524	2	C81367	phosphoenolpyruvat
43	35	45.5	529	2	S12787	potassium channel
44	35	45.5	530	2	JH0167	potassium channel
45	35	45.5	555	1	RGASWA	regulatory protein

ALIGNMENTS

RESULT 1

S48492

DCG1 protein - yeast (*Saccharomyces cerevisiae*)

N;Alternate names: protein YIR030c

C;Species: *Saccharomyces cerevisiae*

C;Date: 02-Dec-1994 #sequence_revision 02-Dec-1994 #text_change 09-Jul-2004

C;Accession: S48492; S19038

R;Rowley, K.

submitted to the EMBL Data Library, October 1994

A;Reference number: S48478

A;Accession: S48492

A;Molecule type: DNA

A;Residues: 1-244 <ROW>

A;Cross-references: UNIPROT:P32460; UNIPARC:UPI0000128FAE; GB:Z47047;

EMBL:Z38061; NID:g603997; PID:g763375; MIPS:YIR030c

R;Yoo, H.S.; Cooper, T.G.

Gene 104, 55-62, 1991

A;Title: Sequences of two adjacent genes, one (DAL2) encoding allantoinase and another (DCG1) sensitive to nitrogen-catabolite repression in *Saccharomyces cerevisiae*.

A;Reference number: JH0442; MUID:92009196; PMID:1916277

A;Accession: S19038

A;Molecule type: DNA

A;Residues: 1-126,'C',128-244 <YOO>

A;Cross-references: UNIPARC:UPI000017923A; GB:M64719

C;Genetics:

A;Gene: SGD:DCG1

A;Cross-references: SGD:S0001469; MIPS:YIR030c

A;Map position: 9R

C;Superfamily: *Saccharomyces cerevisiae* DCG1 protein

C;Keywords: transmembrane protein

F;221-237/Domain: transmembrane #status predicted <TMM>

Query Match 58.4%; Score 45; DB 2; Length 244;
Best Local Similarity 61.5%; Pred. No. 1.6;
Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy 2 QRTSIRATEGCLP 14

| |||:: | |||

Db 51 QETSIKSMEACL P 63

Search completed: June 6, 2006, 05:21:39

Job time : 41 secs

GenCore version 5.1.9

Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: June 6, 2006, 05:12:45 ; Search time 294 Seconds

(without alignments)

47.195 Million cell updates/sec

Title: US-10-655-562A-4
 Perfect score: 77
 Sequence: 1 KQRTSIRATEGCLPS 15

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 2849598 seqs, 925015592 residues

Total number of hits satisfying chosen parameters: 2849598

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : UniProt_7.2:*
 1: uniprot_sprot:*
 2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	DB	ID	Description
1	48	62.3	561	2	Q8SV66_ENCCU	Q8sv66 encephalito
2	46	59.7	105	2	Q4YRP8_PLABE	Q4yrp8 plasmodium
3	45	58.4	86	2	Q3J3C1_RHOS4	Q3j3c1 rhodobacter
4	45	58.4	244	1	DCG1_YEAST	P32460 saccharomyc
5	44	57.1	103	2	Q2QQU0_ORYSA	Q2qqu0 oryza sativ
6	44	57.1	103	2	Q33AN9_ORYSA	Q33an9 oryza sativ
7	42.5	55.2	1184	2	Q57ZH0_9TRYP	Q57zh0 trypanosoma
8	42	54.5	117	2	Q84SP6_ORYSA	Q84sp6 oryza sativ
9	42	54.5	249	2	Q9BPY7_HUMAN	Q9bpy7 homo sapien
10	42	54.5	267	2	Q7L5R2_HUMAN	Q7l5r2 homo sapien
11	42	54.5	267	2	Q9BY14_HUMAN	Q9by14 homo sapien
12	42	54.5	292	2	Q4C9D8_CROWT	Q4c9d8 crocosphaer
13	42	54.5	854	1	ENV_SIVCZ	P17281 chimpanzee
14	42	54.5	987	1	SYV_RHOS4	Q3j4z5 rhodobacter
15	41	53.2	58	2	Q5N9P0_ORYSA	Q5n9p0 oryza sativ
16	41	53.2	238	2	Q3RPZ7_RALME	Q3rpz7 ralstonia m
17	41	53.2	292	2	Q65WI6_MANSM	Q65wi6 mannheimia
18	41	53.2	307	2	Q8KZT6_PSEPU	Q8kzt6 pseudomonas
19	41	53.2	357	2	Q9FFV4_ARATH	Q9ffv4 arabidopsis
20	41	53.2	426	2	Q6NKY2_ARATH	Q6nky2 arabidopsis
21	41	53.2	485	2	Q9SI78_ARATH	Q9si78 arabidopsis
22	41	53.2	494	2	Q2U0Z1_ASPOR	Q2u0z1 aspergillus
23	41	53.2	608	2	Q519L2_ENTHI	Q519l2 entamoeba h
24	40.5	52.6	310	2	Q8U307_PYRFU	Q8u307 pyrococcus
25	40.5	52.6	312	2	O58585_PYRHO	O58585 pyrococcus

26	40	51.9	181	2	Q851P7_ORYSA	Q851p7 oryza sativ
27	40	51.9	213	2	Q5JMM9_ORYSA	Q5jmm9 oryza sativ
28	40	51.9	253	2	Q3H316_9ACTO	Q3h316 nocardioide
29	40	51.9	290	2	Q9PKK6_CHLMU	Q9pkk6 chlamydia m
30	40	51.9	322	2	Q55ZJ6_CRYNE	Q55zj6 cryptococcu
31	40	51.9	322	2	Q5KNW0_CRYNE	Q5knw0 cryptococcu
32	40	51.9	358	2	Q5SV06_MOUSE	Q5sv06 mus musculu
33	40	51.9	368	2	Q550W6_DICDI	Q550w6 dictyosteli
34	40	51.9	368	2	Q86KQ3_DICDI	Q86kq3 dictyosteli
35	40	51.9	539	2	Q2VJ46_9VIRU	Q2vj46 rat adeno-a
36	40	51.9	605	2	Q2VJ47_9VIRU	Q2vj47 rat adeno-a
37	40	51.9	734	2	Q2VJ48_9VIRU	Q2vj48 rat adeno-a
38	40	51.9	759	2	Q4WL45_ASPFU	Q4wl45 aspergillus
39	40	51.9	775	2	Q3JKW4_BURP1	Q3jkw4 burkholderi
40	40	51.9	935	2	Q3EET2_ACTSC	Q3eet2 actinobacil
41	40	51.9	1369	2	Q4Q8A1_LEIMA	Q4q8a1 leishmania
42	40	51.9	23015	2	Q8IQ18_DROME	Q8iq18 drosophila
43	39	50.6	121	2	Q6ZN48_HUMAN	Q6zn48 homo sapien
44	39	50.6	132	2	Q4C816_CROWT	Q4c816 crocosphaer
45	39	50.6	160	1	YFF1_YEAST	P43552 saccharomyc

ALIGNMENTS

RESULT 1

Q8SV66_ENCCU

ID Q8SV66_ENCCU PRELIMINARY; PRT; 561 AA.

AC Q8SV66;

DT 01-JUN-2002, integrated into UniProtKB/TrEMBL.

DT 01-JUN-2002, sequence version 1.

DT 07-FEB-2006, entry version 13.

DE Hypothetical protein ECU06_1590.

GN OrderedLocusNames=ECU06_1590;

OS Encephalitozoon cuniculi.

OC Eukaryota; Fungi; Microsporidia; Unikaryonidae; Encephalitozoon.

OX NCBI_TaxID=6035;

RN [1]

RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].

RC STRAIN=GB-M1;

RX MEDLINE=21576510; PubMed=11719806; DOI=10.1038/35106579;

RA Katinka M.D., Duprat S., Cornillot E., Metenier G., Thomarat F.,

RA Prensier G., Barbe V., Peyretailade E., Brottier P., Wincker P.,

RA Delbac F., El Alaoui H., Peyret P., Saurin W., Gouy M.,

RA Weissenbach J., Vivares C.P.;

RT "Genome sequence and gene compaction of the eukaryote parasite

RT Encephalitozoon cuniculi.";

RL Nature 414:450-453(2001).

CC

CC Copyrighted by the UniProt Consortium, see <http://www.uniprot.org/terms>

CC Distributed under the Creative Commons Attribution-NoDerivs License

CC

DR EMBL; AL590446; CAD25520.1; -; Genomic_DNA.

DR InterPro; IPR002885; PPR.

DR TIGRFAMS; TIGR00756; PPR; 1.

KW Complete proteome; Hypothetical protein.

SQ SEQUENCE 561 AA; 64550 MW; 238E55A1C1C09184 CRC64;

Query Match 62.3%; Score 48; DB 2; Length 561;
Best Local Similarity 53.3%; Pred. No. 5.6;
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 KQRTSIRATEGCLPS 15
|:| :| |||||:
Db 155 KRREMLKAMEGCLPN 169

Search completed: June 6, 2006, 05:20:54
Job time : 297 secs